

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10162012 A

(43) Date of publication of application: 19 . 06 . 98

(51) Int. Cl

G06F 17/30

(21) Application number: 08316667

(71) Applicant: MATSUSHITA ELECTRIC IND CO LTD

(22) Date of filing: 27 . 11 . 96

(72) Inventor: TOKUDA KATSUMI

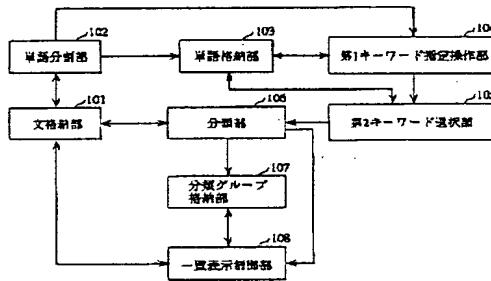
(54) SENTENCE CLASSIFICATION DEVICE

COPYRIGHT: (C)1998,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a sentence classification device that can select a keyword through the use of a word, which a user designates, and can classify sentence set in connection to the designated word through the use of the keyword.

SOLUTION: A word division part 102 word-divides the plural sentences stored in a sentence storage part 101 and stores them in a word storage part 103. A first keyword designation operation part 104 displays the word and the user designates the word. A second keyword selection part 105 selects the word which is the other word in the same sentence as the designated word and which exists in the other sentence as the keyword. A classification part 106 classifies the sentence containing the same keyword into one group, and synthesizes the groups when more than a prescribed rate of the sentences common in the groups exist. A list display control part 108 displays the content of the group with the keyword as a representative word.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-162012

(43)公開日 平成10年(1998)6月19日

(51)Int.Cl. - - - - - 識別記号

G 06 F 17/30

FT

G 06 F 15/401 310D
15/40 370A
15/401 310A
15/403 320D
380E

審査請求 未請求 請求項の数 8 OL (全 11 頁)

(21)出願番号 特願平8-316667

(22)出願日 平成8年(1996)11月27日

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72)発明者 徳田 克己

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

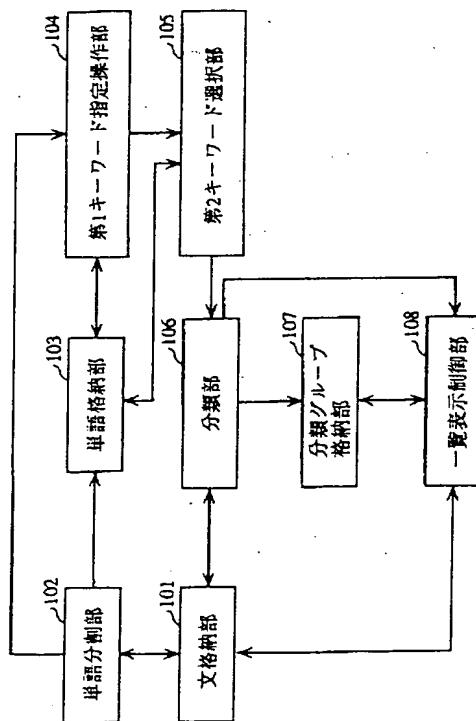
(74)代理人 弁理士 中島 司朗

(54)【発明の名称】 文分類装置

(57)【要約】

【課題】 利用者が指定した単語を用いて、キーワードを選択し、このキーワードを用いて文集合を指定された単語に関連して分類することのできる文分類装置を提供する。

【解決手段】 単語分割部102は、文格納部101に格納されている複数の文を単語分割し、単語格納部103に格納する。第1キーワード指定操作部104は、該単語を表示し、利用者から単語の指定を受ける。第2キーワード選択部105は、指定された単語と同一文中の他の単語で他の文に存在する単語をキーワードとして選択する。分類部106は、同一キーワードを含む文を一つのグループに分類し、グループ間に共通する文が所定割合以上あればグループを統合する。一覧表示制御部108は、キーワードを代表語として、グループの内容を表示する。



【特許請求の範囲】

【請求項1】 複数の文を記憶している文記憶手段と、前記文記憶手段に記憶されている文を単語に分割する単語分割手段と、

前記単語分割手段で分割された単語を表示して前記文記憶手段に記憶されている文を分類するための単語の指定を受け付ける単語指定受付手段と、

前記文記憶手段に記憶されている文の中で前記単語指定受付手段で指定された単語を含む文中の当該単語以外の所定の単語をキーワードとして選択するキーワード選択手段と、

前記キーワード選択手段で選択されたキーワードを代表語として、前記文記憶手段に記憶された文中にキーワードを含む文を同一グループに分類する分類手段と、前記分類手段で分類されたグループの内容を表示させる表示制御手段とを備えることを特徴とする文分類装置。

【請求項2】 前記分類手段でグループに分類されたグループ間で分類された文が所定の割合以上一致するグループ同士を1のグループに統合し、そのグループを代表する代表語として統合前の2つの代表語を連接する統合手段を備え、

前記表示制御手段は統合手段で統合されたグループの内容を表示させる統合表示制御部を有することを特徴とする請求項1記載の文分類装置。

【請求項3】 前記文記憶手段は、複数の文を文番号とともに記憶し、

前記単語分割手段で分割された単語をその文番号とともに記憶している単語記憶手段を備え、

前記キーワード選択手段は、前記単語記憶手段に記憶されている指定された単語と同一文番号の異なる単語が異なる文番号にあるとき、該単語をキーワードとして選択することを特徴とする請求項2記載の文分類装置。

【請求項4】 前記キーワード選択手段は、前記単語記憶手段に記憶されている該単語の出現回数を計数する計数部と、

前記計数部が計数した出現回数の多い単語をキーワードとして選択する出現キーワード選択部とを有することを特徴とする請求項3記載の文分類装置。

【請求項5】 同意語、反意語、類似語の互いに関連する単語を組にして登録している関連単語辞書を備え、

前記単語分割手段は、分割した単語が前記関連単語辞書に登録されているとき、その組にされた単語の1つの単語を前記単語記憶手段にその文番号とともに記憶させる関連単語分割部を有し、

前記分類手段は、前記文記憶手段に記憶された文中にキーワード又は前記関連単語辞書にキーワードと組に登録されている単語を含む文を同一グループに分類することを特徴とする請求項3又は4記載の文分類装置。

【請求項6】 前記分類手段は、代表語と同一グループに分類した文の文番号とを記憶する分類記憶部を有し、

前記統合手段は、

グループ間で分類された文が所定の割合以上一致するか否かを前記分類記憶部に記憶された文番号の一致数から判定する判定部と、

前記判定部が所定の割合以上一致すると判定したとき、前記分類記憶部の2つのグループを、一の連接した代表語と統合した文の文番号とからなる一のグループに更新する更新部とを有することを特徴とする請求項3記載の文分類装置。

【請求項7】 前記表示制御手段は、

前記分類記憶部に記憶された代表語と文番号との数を一覧表示させる一覧表示制御部と、

利用者からの代表語を指定した内容表示の指示を受けると、前記分類記憶部に記憶されている文番号に対応する文を前記文記憶手段から読み出して表示させる内容表示制御部とを有することを特徴とする請求項6記載の文分類装置。

【請求項8】 コンピュータを文分類装置として機能させるため、

20 複数の文を記憶している文記憶手段と、

前記文記憶手段に記憶されている文を単語に分割する単語分割手段と、

前記単語分割手段で分割された単語を表示して前記文記憶手段に記憶されている文を分類するための単語の指定を受け付ける単語指定受付手段と、

前記文記憶手段に記憶されている文の中で前記単語指定受付手段で指定された単語を含む文中の当該単語以外の所定の単語をキーワードとして選択するキーワード選択手段と、

30 前記キーワード選択手段で選択されたキーワードを代表語として、前記文記憶手段に記憶された文中にキーワードを含む文を同一グループに分類する分類手段と、

前記分類手段で分類されたグループの内容を表示させる表示制御手段として機能させるプログラムを記録したことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、複数の文をキーワードを用いて分類する文分類装置に関する。

40 【0002】

【従来の技術】 近年、コンピュータネットワークや文字放送を利用した電子ニュース等の情報自動通信システムが稼働し始めている。この電子ニュース等を効率的に利用するため、記事をキーワードを与えて分類する文分類装置がある。この従来の文分類装置では、利用者が1つのキーワードを与えると、このキーワードを含む文を抽出する。また、複数のキーワードを与えると、それらのキーワードごとにキーワードを含む文に分類する。

【0003】

50 【発明が解決しようとする課題】 ところで、上記装置で

は、与えられたキーワードを含む文を抽出、分類するだけであるので、キーワードに関連する文を広範に分類することが煩雑になる。例えば、新聞記事を参照し、ある業界で主要な地位を占めるA社の動向に注目して、前記業界の動向を調べる場合を考える。

【0004】従来の文分類装置においては、キーワード「A社」を1次検索用のキーワードとして入力する。次に「A社」の動きを表す「DVD開発」や、「輸出不振」というフレーズを検索結果から利用者が見い出して、再度「DVD」、「輸出」などの2次検索用のキーワードとして入力して、「A社」を含む業界の動向を広範囲に検索する。この場合、1次検索の結果が膨大になることが多く、利用者が検索結果を全て参照して2次検索用のキーワードを適切に決定することは非常に困難である。

【0005】また、仮に、2次検索用のキーワードを設定できたとしても、2次検索用のキーワードの組合せによって2次検索の結果は大きく変動する。このため、適切な組合せを求めるために試行錯誤を繰り返す必要がある。本発明は上記欠点に鑑み、利用者が1次検索用のキーワードを指定するだけで、最適な2次検索用のキーワードを自動的に設定し、1次検索用のキーワードを指定するだけで、最適な2次検索用のキーワードを自動的に設定し、1次検索用のキーワードに関連する広範な文の分類ができる文分類装置を提供することを目的とする。

【0006】

【課題を解決するための手段】本発明に係る文分類装置は、複数の文を記憶している文記憶手段と、前記文記憶手段に記憶されている文を単語に分割する単語分割手段と、前記単語分割手段で分割された単語を表示して前記文記憶手段に記憶されている文を分類するための単語の指定を受け付ける単語指定受付手段と、前記文記憶手段に記憶されている文の中で前記単語指定受付手段で指定された単語を含む文中の当該単語以外の所定の単語をキーワードとして選択するキーワード選択手段と、前記キーワード選択手段で選択されたキーワードを代表語として、前記文記憶手段に記憶された文中にキーワードを含む文を同一グループに分類する分類手段と、前記分類手段で分類されたグループの内容を表示させる表示制御手段とを備えることとしている。これによって、一の単語を指定されると、当該単語を含む文から所定のキーワードを選択して、そのキーワードを含む文を同一グループに分類するので、広範な文の分類が可能となる。

【0007】

【発明の実施の形態】以下、本発明に係る文分類装置の実施の形態を図面を用いて説明する。

(実施の形態1) 図1は、本発明に係る文分類装置の実施の形態1の構成図である。この文分類装置は、文格納部101と、単語分割部102と、単語格納部103

と、第1キーワード指定操作部104と、第2キーワード選択部105と、分類部106と、分類グループ格納部107と、一覧表示制御部108とを備えている。

【0008】文格納部101は、磁気ディスク、RAM、RAM等からなり、複数の文をその格納順を示す文番号とともに格納している。この文格納部101には、例えば図2に示すような、文番号201で区分された文字放送の見出し文の内容202が格納されている。単語分割部102は、所定の単語辞書を有し、この文分類装置が起動されると、文格納部101に格納されている文を読み出し、所定の単語に分割し、その文番号とともに単語格納部103に格納する。この際、同一文中に出現する出現回数も記録する。これによって、同一文中の単語を重複して格納することはない。文格納部101に格納されている全ての文について単語分割が終了すると、第1キーワード指定操作部104を起動する。

【0009】ここで、所定の単語辞書とは、助詞等の単語を除いた名詞を中心とした辞書をいう。単語格納部103は、磁気ディスク、RAM等からなり、図3に示すように、単語分割部102で単語に分割された単語301を文格納部101に格納された順番である文番号302とその文中での出現回数303とともに格納している。

【0010】図3は、図2に示した文が文格納部101に格納されていたとき、単語格納部103に格納される単語の一例を示している。第1キーワード指定操作部104は、CRT、液晶ディスプレイ等の表示部を有し、単語分割部102に起動されると、単語格納部103に格納されている単語を読み出し、表示部に表示し、利用者からの第1キーワードの指定を待つ。利用者は、表示部に表示された単語から一つの単語を指定する。第1キーワード指定操作部104は、この指定された単語を第1キーワードとして第2キーワード選択部105に通知する。

【0011】第2キーワード選択部105は、文番号カウンタiと、単語カウンタjとを有し、第1キーワード指定操作部104から第1キーワードの通知を受けると、単語格納部103に格納されている単語から第2キーワードを選択し、選択した第2キーワードを分類部106に通知する。第1キーワードの通知を受けると、文番号カウンタiに「1」を設定する。また文番号カウンタiは、単語カウンタjの値が文番号iの文の単語数NWを超えたとき、又は文番号iの文が第1キーワードを含まないとき、「1」を加えられる。

【0012】次に、文番号iの文の全ての単語を単語格納部103から読み出し、この単語の中に第1キーワード指定操作部104から通知された第1キーワードに一致する単語があるか否かを判定し、あるときは、この文番号iの文の単語の総数NWを読み出し、単語カウンタjに「1」を設定する。単語カウンタjは、文番号iの

文の j 番目の単語 $W_{i,j}$ が第 1 キーワードであるとき、又は単語 $W_{i,j}$ が単語格納部 103 の文番号 i 以外の文番号の単語に一致するものがないとき、「1」を加えられる。

【0013】また、単語格納部 103 から文番号 i の j 番目の単語 $W_{i,j}$ を読み出し、単語 $W_{i,j}$ が第 1 キーワードと一致するか否かを判定し、一致しないときは、単語格納部 103 に格納されている文番号 i 以外の文の単語に一致するものがあるか否かを判定する。一致するものがあるときは、この単語 $W_{i,j}$ を第 2 キーワードとして選択し、分類部 106 に通知する。更に、文番号カウンタ i の値が単語格納部 103 に格納されている文の総数（文番号の最大） N_S を超えたか否かを判定し、超えているときは、分類部 106 を起動する。

【0014】図 3 に示す単語群が単語格納部 103 に格納されているとき、第 1 キーワード指定操作部 104 で単語「A社」が第 1 キーワードとして指定されると、第 2 キーワード選択部 105 は、第 1 キーワード「A社」と同一の文番号「1」及び「5」を持つ単語から文番号「1」又は「5」以外の文番号を持つ単語群に一致する単語「年度」、「決算」、「発表」、「円」、「新型」、「発売」を第 2 キーワードとして選択し、分類部 106 に通知する。

【0015】分類部 106 は、RAM 等からなる第 2 キーワード記憶部（図示せず）と、文番号カウンタ i と、第 2 キーワードカウンタ k と、グループカウンタ m と、統合カウンタ n_e と、グループ比較カウンタ n を有する。第 2 キーワード選択部 105 から第 2 キーワードを通知されると順次第 2 キーワード記憶部に記憶させ、また起動されると、文格納部 101 に記憶されている文の総数 N_S を読み出し、文番号カウンタ i に「1」を設定する。文番号カウンタ i は、第 2 キーワードカウンタ k の値が第 2 キーワード記憶部に記憶されている第 2 キーワードの総数 $N_K 2$ の値を超えたとき、「1」を加えられる。

【0016】分類部 106 は、文番号カウンタ i に「1」が設定されたとき、またはインクリメントされたとき、文格納部 101 に格納されている文番号 i の文 S_i を読み込む。続いて、第 2 キーワード記憶部に記憶されている第 2 キーワードの総数 $N_K 2$ を読み出し、第 2 キーワードカウンタ k に「1」を設定する。第 2 キーワードカウンタ k は、文格納部 101 に格納されている文 S_i に第 2 キーワード記憶部の k 番目に記憶されている第 2 キーワードが含まれているか否かを判定した後に「1」を加えられる。

【0017】分類部 106 は、第 2 キーワードカウンタ k に「1」を設定したとき、又は第 2 キーワードカウンタ k の値が第 2 キーワードの総数 $N_K 2$ を超えないとき、第 2 キーワード記憶部に k 番目に記憶された第 2 キーワード $K 2 k$ を読み出し、文 S_i が第 2 キー

ワードを含んでいるか否かを判定する。文 S_i に第 2 キーワードが含まれているときは、第 2 キーワード $K 2 k$ を分類グループ代表語として分類グループ G_k に文 S_i を分類する。分類結果を分類グループ格納部 107 に分類グループ代表語と、文番号とを対応して格納するとともに、その分類グループに含まれる文の数を格納する。なお、既に分類グループ代表語が格納されているときには、文番号を対応して格納して、文の数を「1」増やす。分類部 106 が第 2 キーワード選択部 105 から

10 第 2 キーワード「年度」、「決算」、「発表」、「円」、「新型」、「発売」を通知され、図 2 に示した文が文格納部 101 に格納されているとき、分類グループ格納部 107 には、図 4 に示すように、分類グループ代表語（第 2 キーワード）ごとに各文が分類される。即ち、分類グループ代表語 401 の「年度」を含む文番号 402 の文「1」、「2」、「3」が 1 つの分類グループ 403 として分類される。この分類グループ 403 には、文の数 404 が「3」であることが示されている。また、図 4 からわかるように、分類部 106 では、最初、文格納部 101 に格納されている各文を第 2 キーワードの数と同じ分類グループ数として分類する。

【0018】分類部 106 は、文格納部 101 に格納されている全ての文について、第 2 キーワードを分類グループ代表語とする分類を終了すると、分類グループ格納部 107 に格納した分類グループ数 $N_K 2$ を読み出し、変数 n_g に代入する。また、変数 n_g が分類グループの統合前の初期分類グループ数 n_{g0} と一致しないとき、変数 n_g に変数 n_g を代入する。ここで、変数 n_g は、現時点での分類グループ数を示している。

30 【0019】次に、初期分類グループ数 n_{g0} に変数 n_g を代入し、グループカウンタ m に「1」を設定する。ここで、グループカウンタ m の値は、分類グループ格納部 107 に格納した分類グループの順番を示す。グループカウンタ m に「1」を設定したとき、又はグループカウンタ m の値が変数 n_g の値を超えないとき、統合カウンタ n_e に「0」を設定し、グループ比較カウンタ n に「 $m+1$ 」を設定する。ここで、グループ比較カウンタ n の値は、分類グループ格納部 107 に格納した分類グループの順番を示す。

40 【0020】分類部 106 は、分類グループ格納部 107 の i 番目と j 番目に格納されている分類グループ G_i と G_j とに分類された文の所定の割合、例えば 75% 以上が同一文か否かを判定する。即ち、分類グループ代表語に対応して記憶されている文番号が 75% 以上共通しているか否かを判定する。否と判定したときは、統合カウンタ n に「1」を加え、 n の値が変数 n_g の値を超えるか否かを判定する。変数 n_g の値を超えるときは、変数 n_g に変数 n_g の値から統合カウンタ n_e の値を減算した値を代入する。

50 【0021】次に、グループカウンタ m に「1」を加え

て、 m の値が変数 ng を超えるか否かを判定し、超えるときは変数 ng が初期分類グループ数 ng_0 に一致するか否かを判定し、一致するときは、一覧表示制御部108を起動する。文の75%以上が同一文であると判定したときは、分類グループ G_m に分類グループ G_n を統合する。即ち、分類グループ格納部に格納されている2つの分類グループ G_m と G_n とのそれぞれの分類グループ代表語を連接して1つの分類グループの分類グループ代表語とし、対応して文番号と文の数とを格納する。次に、統合カウンタ ne に「1」を加えて、上述の否と判定したときと同様の処理をする。

【0022】以上のような処理の結果、分類グループ格納部107には、図4に示した分類グループ403、405が統合され、図5に示すように分類グループ501となる。即ち、分類グループ403の文番号「1」、「2」、「3」と分類グループ405の文番号「1」、「2」、「3」、「4」とが75%以上一致しているからである。

【0023】同様に、図4に示す分類グループ406、407は、文番号が100%一致しているので、図5に示す分類グループ502に統合される。なお、分類グループ代表語401は、分類グループ501では「年度・決算」に、分類グループ502では「新型・発売」にそれぞれ連接されている。一覧表示制御部108は、液晶ディスプレイ、CRT等からなる表示部を有し、分類部106に起動されると、分類グループ格納部107に格納されている分類グループの内容を読み出し、図6に示すように表示部に表示させる。

【0024】表示内容は、分類グループ代表語601とその分類グループに含まれる文の数602との一覧である。ここで、図示しない操作部からカーソルキーによりカーソル603を移動して、分類グループ代表語「新型・発売」を項目選択して、リターンキーにより内容表示に切り換えると、図7に示すように分類グループ502に分類された3つの文が表示される。

【0025】この際、一覧表示制御部108は、分類グループ格納部107に格納されている文番号を読み出し、文格納部101に格納されている対応する文を読み出して表示させる。次に、本実施の形態の動作を図8、図9、図10のフローチャートを用いて説明する。

【0026】先ず、文分類装置が起動されると、単語分割部102は、単語辞書を用いて文格納部101に格納されている全ての文を単語に分割して単語格納部103に格納する(S802)。第1キーワード指定操作部104は、利用者から第1キーワードの指定を受け付ける(S804)。

【0027】第2キーワード選択部105は、単語格納部103に格納されている文の数 NS を読み出して記憶するとともに文番号カウンタ i に「1」を設定する(S806)。単語格納部103の文番号 i の全ての単語 W

を読み込み(S808)、単語 W の中に第1キーワードに一致する単語が有るか否かを判定する(S810)。【0028】一致する単語がなければS824に移り、一致する単語が有れば、文番号 i の単語数 NW を読み出して記憶するとともに単語カウンタ j に「1」を設定する(S812)。文番号 i の j 番目の単語 $W_{i,j}$ は第1キーワードと一致するか否かを判定し(S814)、一致するときはS820に移り、一致しないときは、単語 $W_{i,j}$ と単語格納部103に格納されている文番号 i 以外の単語に一致する単語が有るか否かを判定する(S816)。

【0029】有るときは、単語 $W_{i,j}$ を第2キーワードの1つとして分類部106に通知し(S818)、ないときはS820に移る。S820において、単語カウンタ j に「1」を加え、単語カウンタ j の値が文番号 i の単語数 NW を超えるか否か判定し(S822)、超えないときはS814に戻り、超えるときはS824に移る。

【0030】S824において、文番号カウンタ i に「1」を加え、文番号カウンタ i の値が単語格納部103に格納されている文の総数 NS を超えるか否かを判定し(S826)、超えないときはS808に戻り、超えるときはS902に移る。分類部106は、S902において、文格納部101に格納されている文の総数 NS を読み出して記憶し、文番号カウンタ i に「1」を設定する。次に文格納部101に格納されている文番号 i の文 S_i を読み込み(S904)、第2キーワードの数 NK_2 として、第2キーワードカウンタ k に「1」を設定する(S906)。

【0031】次に、文 S_i に k 番目の第2キーワード K_{2k} を含むか否かを判定し(S908)、含むときには文 S_i を第2キーワード K_{2k} を分類グループ代表語とする分類グループ G_k に分類し、分類グループ格納部107にその文番号 i を格納する(S910)。その後、又はS908において含まないと判定したとき、第2キーワードカウンタ k に「1」を加え(S912)、第2キーワードカウンタ k の値が第2キーワードの数 NK_2 を超えたか否かを判定し(S914)、超えていないときはS908に戻り、超えたときは文番号カウンタ i に「1」を加え(S916)、文番号カウンタ i の値が文の総数 NS を超えたか否かを判定する(S918)。超えていないときはS904に戻り、超えているときは、第1回目の分類を終了し、現在の分類グループ数 ng を読み出す。ここでは $ng = NK_2$ である(S1002)。

【0032】次に初期分類グループ数 ng_0 に「 ng 」を代入し、グループカウンタ m に「1」を設定する(S1004)。統合カウンタ ne に「0」を設定し、グループ比較カウンタ n に「 $m+1$ 」を設定する(S1006)。分類グループ格納部107に格納されている分類

グループGmと分類グループGnとに分類された文の75%以上が同一文か否かを判定する(S1008)。肯定のときは、分類グループGmに分類グループGnを統合し(S1010)、否であればS1014に移る。分類グループ数が「1」減ったので統合カウンタneに「1」を加え(S1012)、グループ比較カウンタnに「1」を加え(S1014)、グループ比較カウンタnの値が分類グループ数ngを超えているか否か判定する(S1016)。超えていればS1018に移り、否のときはS1008に戻る。

【0033】次に、現時点での初期分類グループ数ngに「ng-ne」の値を代入し(S1018)、グループカウンタmに「1」を加え(S1020)、mの値がngを超えるか否かを判定する(S1022)。超えなければS1006に戻り、超えていればngとng0とが一致するか否かを判定し(S1024)、否であればS1004に戻り、一致すればS1026に移る。

【0034】S1026において、一覧表示制御部108は、分類グループ格納部107に格納されている分類グループごとに分類グループ代表語と分類グループに属する文の数とを対応づけて一覧表示し、利用者から指示があれば、その文の内容を表示して(S1026)、処理を終了する。なお、上記実施の形態では、分類グループ統合のための文の一一致度を75%としたけれども、文格納部101に格納されている文の種類によって適当な値に変更することも可能である。また、第2キーワードを選択する際に、単語格納部103において2回以上出現したものに限ったけれども、この出現回数を1回を含む任意の回数に変更することもできる。更に、第2キーワードとしてふさわしくない単語を集めた無効キーワード辞書を備えるようにして、第2キーワード選択部105が選択した第2キーワードであって無効キーワード辞書に一致する単語があるときには、除外するようにして明らかに不要な単語が第2キーワードにならないようにすることができる。

(実施の形態2) 図11は、本発明に係る文分類装置の実施の形態2の構成図である。

【0035】この文分類装置は、文格納部101と、代表単語辞書1101と、単語分割部1102と、単語格納部103と、第1キーワード指定操作部104と、第2キーワード選択部1103と、分類部1104と、分類グループ格納部107と、一覧表示制御部108とを備えている。なお、上記実施の形態1と同一の構成部分には同一の符号を付しその説明を省略し、本実施の形態固有の構成部分について説明する。

【0036】代表単語辞書1101は、図12に示すように、代表単語1201と、代表単語と関連の深い単語、同意語、類似語、反意語を組にした単語組1202とを対応して登録している。即ち、「利益」と「増益」とは関連の深い単語であり、「増益」と「減益」とは反

意語であり、これらの単語を「利益」を代表単語として登録している。

【0037】単語分割部1102は、上記実施の形態1の単語分割部102の機能に加えて、文格納部101に格納されている文を単語単位に分割するに際して、分割した単語が代表単語辞書1101の単語組1202にあるときには、その代表単語1201を単語格納部103に格納する。このように、関連する単語を一まとめにして、後にする第2キーワードに適切な単語を選択する。

10 【0038】第2キーワード選択部1103は、上記第2キーワード選択部105の機能に加えて、第2キーワードとして選択した単語の単語格納部103での出現回数を計数する計数部を有する。そして、この出現回数の多いものから順に一定個数の単語を第2キーワードとして選択し、分類部1104に通知する。このように出現回数の多い単語のみを第2キーワードとすることによって、分類グループの数を減少しつつも、関連ある文をできるだけ多く分類することができる。

【0039】分類部1104は、上記分類部106の機能に加えて、文格納部101に格納されている文を読み出して、その構成単語に第2キーワードを含んでいるか否かを調べる際に、代表単語辞書1101の単語組を含んでいるときにも、対応する代表単語の第2キーワードを含んでいるものとする。このようにすることによって、より広範に文を分類グループに分類することができる。

【0040】次に本実施の形態の動作を説明する。上記実施の形態1の動作と殆ど異なるところがないので、図8、図9、図10のフローチャートの異なる部分のみ説明する。S802に替えて、単語分類部1102は、文格納部101に格納されている全ての文を単語に分割して、該単語が代表単語辞書1101の単語組に登録されているか否かを調べ、登録されていないときはそのまま、登録されているときは対応する代表単語に換えて単語格納部103に格納する。

【0041】S818において、第2キーワード選択部1103は、単語Wijを第2キーワードの候補単語として記憶しておき、S826とS902との間に、候補単語として記憶している単語の出現回数を単語格納部103に格納されている該単語の出現回数を加算して求め、出現回数の多いものから、所定数の候補単語を第2キーワードとして選択して、分類部1104に通知する。

【0042】S908において、分類部1104は、文Siにk番目の第2キーワードK2k又は代表単語辞書1101の代表単語である第2キーワードK2kの対応する単語組の単語を含むか否かを判定する。なお、本実施の形態において、代表単語辞書1101には、予め代表単語が登録されていなければ、変形例として、単語組のみを登録しておいて、単語組の中で文格納部101

に格納された文中の出現回数の多い単語を代表単語とするようにしてもよい。

【0043】なお、上記実施の形態1、2では、図1、図11に示したような構成で本発明に係る文分類装置を実現したけれども、本発明はプログラムによって実現し、それをフロッピーディスク等の記録媒体に記録して移送することにより、他のコンピュータ・システムで容易に実施することができる。図13は、これをフロッピーディスクで実施する場合の説明図である。

【0044】記録媒体本体であるフロッピーディスク1301の物理フォーマットは、同心円上に外周から内周に向かってトラック1、2、…、80を作成し、角度方向に1.6のセクタに分割している。このように割り当てられた領域に従って、プログラムを記録する。このフロッピーディスク1301は、ケース1302に収納され、これによって、ディスクを埃や外部からの衝撃から守り、安全に移送することができる。

【0045】図14は、フロッピーディスク1301にプログラムの記録再生を行うことを説明する図である。図示のようにコンピュータ・システム1401にフロッピーディスクドライブ1402を接続することにより、ディスク1301に対してプログラムを記録再生することができる。ディスク1301はフロッピーディスクドライブ1402に、挿入口1403を介して組込み、および取り出しがなされる。記録する場合はコンピュータ・システム1401からプログラムをフロッピーディスクドライブ1402によってディスク1301に記録する。再生する場合は、フロッピーディスクドライブ1402がプログラムをディスク1301から読み出し、コンピュータ・システム1401に転送する。

【0046】なお、この実施の形態においては、記録媒体としてフロッピーディスクを用いて説明を行ったが、光ディスクを用いても同様に行うことができる。また記録媒体はこれらに限られず、ICカード、ROMカセット等、プログラムを記録できるものであれば、同様に実施することができる。

【0047】

【発明の効果】以上説明したように、本発明は、複数の文を記憶している文記憶手段と、前記文記憶手段に記憶されている文を単語に分割する単語分割手段と、前記単語分割手段で分割された単語を表示して前記文記憶手段に記憶されている文を分類するための単語の指定を受け付ける単語指定受付手段と、前記文記憶手段に記憶されている文の中で前記単語指定受付手段で指定された単語を含む文中の当該単語以外の所定の単語をキーワードとして選択するキーワード選択手段と、前記キーワード選択手段で選択されたキーワードを代表語として、前記文記憶手段に記憶された文中にキーワードを含む文を同一グループに分類する分類手段と、前記分類手段で分類されたグループの内容を表示させる表示制御手段とを備え

ることとしているので、利用者は1つの単語を指定するだけで指定した単語に関連する文記憶手段に記憶されている文を広範に分類した分類結果を見ることができる。

【0048】また、前記分類手段でグループに分類されたグループ間で分類された文が所定の割合以上一致するグループ同士を1のグループに統合し、そのグループを代表する代表語として統合前の2つの代表語を接続する統合手段を備え、前記表示制御手段は統合手段で統合されたグループの内容を表示させる統合表示制御部を有することとしているので、利用者は、より集約された状態で分類されたグループの内容を見ることができる。

【0049】また、前記文記憶手段は、複数の文を文番号とともに記憶し、前記単語分割手段で分割された単語をその文番号とともに記憶している単語記憶手段を備え、前記キーワード選択手段は、前記単語記憶手段に記憶されている指定された単語と同一文番号の異なる単語が異なる文番号にあるとき、該単語をキーワードとして選択することとしているので、キーワードの選択を容易にすることができる。

【0050】また、前記キーワード選択手段は、前記単語記憶手段に記憶されている該単語の出現回数を計数する計数部と、前記計数部が計数した出現回数の多い単語をキーワードとして選択する出現キーワード選択部とを有することとしているので、より重要なキーワードを選択して文を分類することができ、かつ、分類に無駄がなくなる。

【0051】また、同意語、反意語、類似語の互いに関連する単語を組にして登録している関連単語辞書を備え、前記単語分割手段は、分割した単語が前記関連単語

辞書に登録されているとき、その組にされた単語の1つの単語を前記単語記憶手段にその文番号とともに記憶させる関連単語分割部を有し、前記分類手段は、前記文記憶手段に記憶された文中にキーワード又は前記関連単語辞書にキーワードと組に登録されている単語を含む文を同一グループに分類することとしているので、キーワードに関連する単語の重複がなくなり、分類効率が向上する。

【0052】また、前記分類手段は、代表語と同一グループに分類した文の文番号とを記憶する分類記憶部を有し、前記統合手段は、グループ間で分類された文が所定の割合以上一致するか否かを前記分類記憶部に記憶された文番号の一致数から判定する判定部と、前記判定部が所定の割合以上一致すると判定したとき、前記分類記憶部の2つのグループを、一の接続した代表語と統合した文の文番号とからなる一のグループに更新する更新部とを有することとしているので、グループの統合を容易にすることができる。

【0053】また、前記表示制御手段は、前記分類記憶部に記憶された代表語と文番号との数を一覧表示させる一覧表示制御部と、利用者からの代表語を指定した内容

表示の指示を受けると、前記分類記憶部に記憶されている文番号に対応する文を前記文記憶手段から読み出して表示させる内容表示制御部とを有することとしているので、利用者は、分類された文の内容を容易に見ることができ、文分類の利用効率が向上する。

【0054】更に、コンピュータを文分類装置として機能させるため、複数の文を記憶している文記憶手段と、前記文記憶手段に記憶されている文を単語に分割する単語分割手段と、前記単語分割手段で分割された単語を表示して前記文記憶手段に記憶されている文を分類するための単語の指定を受け付ける単語指定受付手段と、前記文記憶手段に記憶されている文の中で前記単語指定受付手段で指定された単語を含む文中の当該単語以外の所定の単語をキーワードとして選択するキーワード選択手段と、前記キーワード選択手段で選択されたキーワードを代表語として、前記文記憶手段に記憶された文中にキーワードを含む文を同一グループに分類する分類手段と、前記分類手段で分類されたグループの内容を表示させる表示制御手段として機能させるプログラムを記録した記録媒体としているので、文分類機能を有しない他のパソコン用コンピュータと相俟って上記文分類装置と同様の効果が得られる。

【図面の簡単な説明】

【図1】本発明に係る文分類装置の実施の形態1の構成図である。

【図2】上記実施の形態の文格納部に格納されている文の一例を示す図である。

【図3】上記実施の形態の単語格納部に格納された単語の一例を示す図である。

【図4】上記実施の形態の分類グループ格納部に格納された内容の一例を示す図である。

【図5】上記実施の形態の分類グループ格納部に格納さ

れた内容の一例を示す図である。

【図6】上記実施の形態の一覧表示制御部によって表示された内容を示す図である。

【図7】上記図6の表示内容の変更された様子を示す図である。

【図8】上記実施の形態の動作を説明するフローチャートである。

【図9】上記実施の形態の動作を説明するフローチャートである。

10 【図10】上記実施の形態の動作を説明するフローチャートである。

【図11】本発明に係る文分類装置の実施の形態2の構成図である。

【図12】上記実施の形態の代表単語辞書の内容を示す図である。

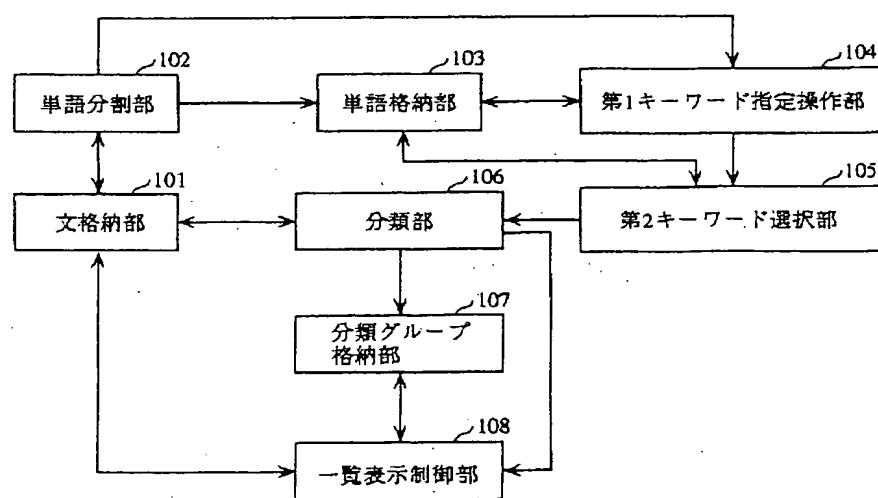
【図13】上記実施の形態1又は2で説明した文分類装置の各構成要素の機能を実施するプログラムを記録した記録媒体の説明図である。

20 【図14】上記記録媒体のコンピュータシステムへの装着を説明する図である。

【符号の説明】

- 101 文格納部
- 102、1102 単語分割部
- 103 単語格納部
- 104 第1キーワード指定操作部
- 105、1103 第2キーワード選択部
- 106、1104 分類部
- 107 分類グループ格納部
- 108 一覧表示制御部
- 1101 代表単語辞書
- 1301 フロッピーディスク
- 1401 コンピュータ・システム

【図1】



【図12】

代表単語	単語組
利益	増益、減益
製品	商品
新型	最新型、改良型
:	:

【図2】

文番号	内容
1	A社は96年度第3四半期の決算を発表、30億円大幅増益
2	B社は95年度の決算を発表、增收減益に
3	C社は本年度の決算予想を上方修正、10億円の増益要因
4	B社は不動産を売却、決算黒字は維持
5	A社は新型冷蔵庫を1月に発売
6	B社は新型エアコンを発売
7	C社は新型ソフトを発売

【図3】

文番号	単語	出現回数
1	A社	1
1	96	1
1	年度	1
1	第3	1
1	四半期	1
1	決算	1
1	発表	1
1	30億	1
1	円	1
1	大幅	1
1	増益	1
2	B社	1
2	95	1
2	年度	1
⋮	⋮	⋮

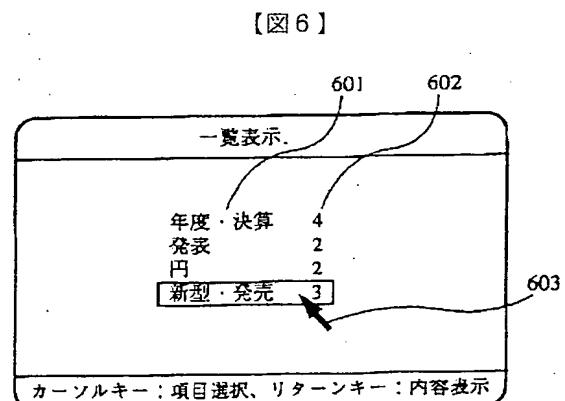
【図4】

分類グループ	代表語	文数	文番号
年度		3	1,2,3
決算		4	1,2,3,4
発表		2	1,2
円		2	1,3
新型		3	5,6,7
発売		3	5,6,7

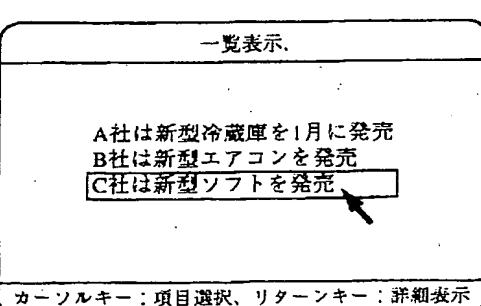
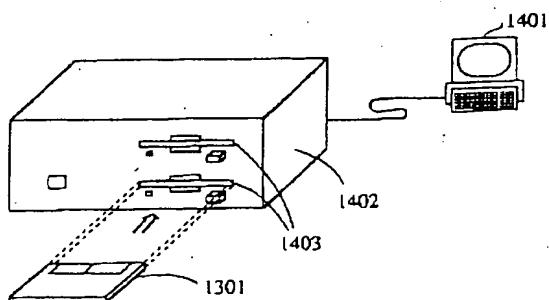
【図5】

分類グループ	代表語	文数	文番号
年度・決算		4	1,2,3,4
発表		2	1,2
円		2	1,3
新型・発売		3	5,6,7

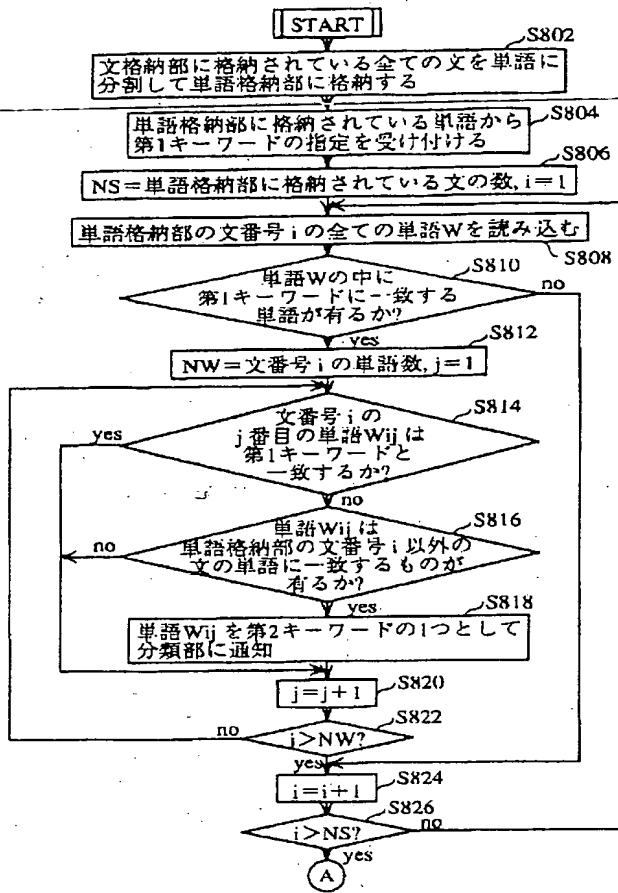
【図7】



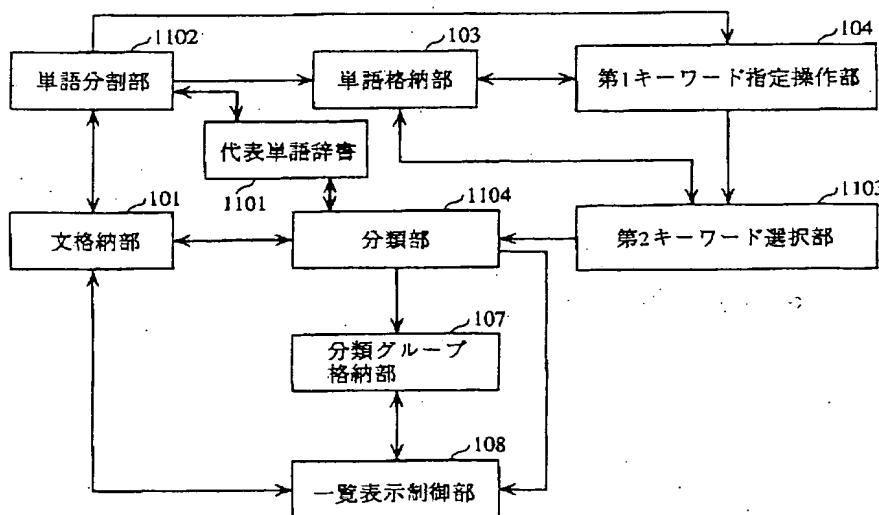
【図14】



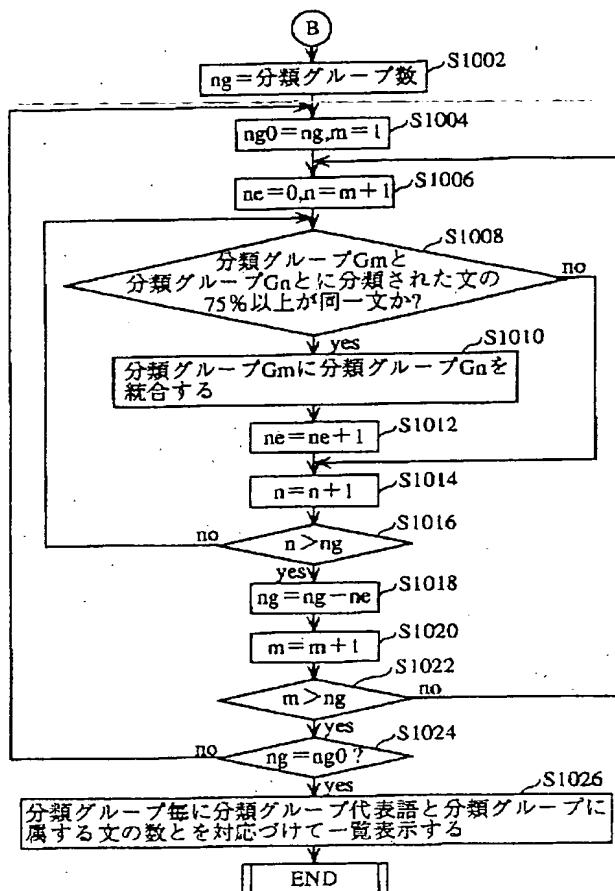
【図8】



【図11】



【図10】



【図13】

